An Agent-Based Approach for Modeling Population Behavior and Health with Application to Tobacco Use

Stephen J. Verzi[1], Benjamin Apelberg[2], Brian Rostron[2], Nancy S. Brodsky[1], Theresa J. Brown[1]

[1] Sandia National Laboratories, Albuquerque, New Mexico, USA

[2] Food & Drug Administration, Rockville, Maryland, USA

Corresponding author: sjverzi@sandia.gov

Abstract

Background. Population dynamics modeling can be an effective tool for forecasting population health effects and identifying sound policy decisions. Systems dynamics modeling of population groups through stocks and flows is highly effective for this purpose, but representing diverse population characteristics is organizationally difficult. Agent-based modeling simplifies this process by managing information at the individual level. This paper presents a Population Structure Model (PSM) for evaluating the population impacts of changes in individual behavior using cigarette smoking as the motivating example.

Methods. Our model incorporates agent-based, discrete-event, dynamical systems and microsimulation to project population health outcomes. In this analysis, we run 100 simulations of a 1:10,000 scale US population (28,142 agents in year 2000) using a discrete-event Markov chain approach to describe state transitions for initiating and quitting smoking and dying. New agents are added through migration and birth. Model inputs are obtained from US Census, vital statistics, and national health survey data.

Results. The PSM population projections are consistent with US Census projections. With constant initiation and cessation rates, adult smoking prevalence is projected to decline from about 18.6% in 2010 to around 12.5% in 2050. Projected smoking prevalence is consistent with observed data from 2000-2011 and projections from other simulation models. We also estimate 401,000 (95% CI: 389,000-413,000) smoking-attributable deaths in 2000 among individuals ages 35+, consistent with estimates from the Centers for Disease Control and Prevention (393,000 deaths, excluding fires and secondhand smoke deaths).

Conclusions. The advantages of our approach include: 1) detailed individual-level characterization of health behaviors and outcomes; 2) extensive flexibility in representing demographic diversity; 3) use of empirical data in model inputs; and 4) capacity for aggregation and analysis at various levels of population size. This

approach allows for the effective prioritization of policy options through evaluation of their short- and long-term health impacts.

Introduction

Problem Definition and Objective

The need for prioritizing scarce public health resources and planning for future health care needs necessitates development of methods for predicting how changes in individual behaviors and associated disease risks will translate into long-term health trends in a population. Although the effectiveness of a policy intervention can be assessed through empirical evidence, evaluation of a policy through data collection and monitoring can be time consuming and expensive. In addition, attribution of the effect a particular policy or intervention to public health outcomes can be methodologically challenging, particularly as multiple interventions are implemented over the same period of time. Computer modeling and simulation provide a means for estimating and comparing the probable effects of different public health strategies prior to implementation. Such simulation can lead to better initial choices and reduce the time and expense required to identify effective strategies.

Population dynamics modeling is a well-established research area that can provide insight into future health impacts associated with current and future risk behaviors [1]-[3]. Population dynamics modeling has been a major focus of mathematical biology for well over 200 years [4], [5], representing population changes over time due to disease and death as well as changes in the environment and available resources. A mathematical model for population dynamics should represent the significant processes affecting birth, death, and migration [1], [6]. More sophisticated models of human population dynamics incorporate behavioral diversity and its resultant consequences. Unfortunately, many human population models do not include representation of human behavior, and those that do typically address the issue at the population or sub-population level, rather than at the individual level.

Integration of agent-based modeling with population dynamics modeling allows direct representation of the behaviors of individuals. Behaviors can be represented using Markov chains that allow the parameterization of the probability of each type of behavior for each individual. An agent's behavior affects its future health, and thereby the health of the total population. In this paper we present a population structure dynamics modeling framework that incorporates individual agent decisions, thereby offering extensive flexibility in representing individual and population health outcomes. We use individuals' cigarette smoking behaviors as our motivating example, and measure consequences of these individual behaviors across the entire population.

We begin by presenting a brief overview of existing relevant computational models and how they have been used to model population dynamics. We then present a

detailed description of our newly proposed model and its mathematical operation. In the following section, we then demonstrate the use of our model to assess the population-level effects of individuals' likelihoods of abstaining from, initiating or quitting cigarette smoking. Finally, we present conclusions and directions for future research.

Background

Most early population models used time-based update equations, in which the birth rate (br) adds to and the death rate (dr) subtracts from the population at each point in time [1], [7].

$$population^{t+1} = population^{t} + br * population^{t} - dr * population^{t} \quad (1)$$

This equation can be restated in difference form or even differential form, under assumptions of differentiability for population with respect to time. Also note that incorporating migration requires one or two additional rates (one for net migration or two for immigration and emigration, respectively) that affect population change over time. An important element missing from the basic population update equation above is the dependence of the death rate on risk behavior. The equation therefore cannot represent individual behaviors and their consequences. However, various approaches can be used to model population behavior, risks, and health. Several commonly used techniques are discussed below.

Discrete-Event Methods

Discrete-event based simulations are useful for representing events occurring at particular points in time [6]-[8], where the timestep can be a fixed interval or driven by the occurrence of events. Discrete event models represent the chronology of changes in state using a list of times associated with events. Typically many simulation runs are aggregated to determine mean values for model outputs.

Continuous Methods

Simulations that model continuous change are well represented using a series of differential equations. In these models, determining the precise mathematical form of these equations must be done *a priori*. Model outputs from these simulations typically will be smooth, but they can include (continuous) stochastic variables. One example is the Lotka-Volterra (or predator-prey) equations describing the continuous (differentiable) relationship between two populations, one of which preys upon the other [9].

System Dynamics

System dynamics (SD) simulations model the temporal operation of an entire system and its sub-systems, typically using a set of coupled, non-linear differential equations [6], [10], [11]. The equations mathematically represent the flow of resource or stock from one part of the system or sub-system to another.

Population dynamics can be very effectively represented using differential equations, especially for population-level birth, death and immigration, and thus are very well suited for SD methods [6], [8], [11]. SD models have also been used to characterize the population health effects of cigarette smoking and various policy options [12], [13].

In an SD model, individuals are not typically represented independent from one-another. Instead, groups of like individuals are shown as stocks of populations (or sub-populations). Behaviors may be incorporated into an SD model by use of additional stocks and flows representing behavioral states and rates of state transitions, respectively. SD modeling of populations with very diverse and/or overlapping demographics and behaviors can become organizationally complicated, requiring many stocks and/or co-flows. As an example, two demographic categories (male and female) and three behaviors (never smoker, current smoker and former smoker) require six stocks, one for each smoker category of each sex. To include more diverse demographics (such as individual age, race, and ethnicity) and different disease states, a stock is needed for each unique combination of states, and flows are needed between stocks. Overlapping characteristics add further complication. A pure SD approach becomes highly complicated, and flexibility is lost as changes to the demographic or disease states of interest require a restructuring of the model.

Agent-based Modeling

Agent-based simulations model behavior and communication among individuals [6], [7], [14]-[16]. Typically each agent is autonomous and sometimes independent of other agents (i.e., no inter-agent communication). In agent-based models, agent states of various sub-populations can be measured and aggregated across many simulation runs at each point in time in the simulation. Demographic and behavioral information can be stored at the agent level and agents can be aggregated for analysis according to any relevant characteristic (e.g., age, disease-type, etc.) without changing model structure. Agent-based modeling therefore facilitates analysis of complicated demographics, behaviors, and health consequences through management of information at the level of the agent. This method has the advantage of flexibility in terms of representational capability, but can require large numbers of agents and multiple simulation runs to assess the stochasticity inherent in the agents or their individual behaviors.

Microsimulation

Microsimulation, also called microanalytic simulation, models represent individuals' opinions or behavior based on real-world data [6], [17], [18]. Microsimulation often uses data that reflect the behavioral tendencies or preferences of individuals in the population being modeled. These models are typically extremely detailed and data intensive. Microsimulation has been used to model a variety of health-related topics including health care utilization and expenditures, infectious disease transmission and prevention, and the health and economic effects of disease screening [19].

Warner et al., for example, used microsimulation to model the health and economic impacts of a workplace smoking cessation program [20].

Model Use Case: Health Impacts of Tobacco Use

Various models presented in the research literature focus on the health consequences of smoking. We discuss three of the more prominent of these models and their contributions to the field below. Additional models have been developed by Richardson [21], Roberts [22], Cavana and Tobias [23], Tengs/Osgood/Lin/Chen [24], [25], and Apelberg [26].

SimSmoke

SimSmoke was developed in MS Excel by David Levy and collaborators and contains a demographic population dynamics model that uses difference equations to update the total population over time [27]-[29]. It is used to track the flow of persons between population sub-groups, such as never, former, and current smokers, as well as mortality. SimSmoke tracks groups by age, sex, race/ethnicity and smoking status [27]-[29]. SimSmoke uses US Census and national health survey data as inputs and can analyze a detailed set of cessation strategies. The model has been validated in U.S. and non-U.S. populations [30]-[33].

Mendez and Warner Models

Mendez and Warner have published extensively on modeling cigarette smoking use [34]-[36]. Their population dynamics models comprise a series of dynamic equations parameterized by cessation. Mendez extends the earlier model using system dynamics to provide a detailed representation of youth experimentation as well as relative risks that vary by sex, age, age at cessation, and years since quitting for smokers [37].

The Mendez-Warner models use birth cohorts and do not include migration effects. Initiation in all models occurs entirely at 18 years of age, using an aggregate initiation rate derived from a target prevalence value.

Prevention Impacts Simulation Model (PRISM)

The Prevention Impacts Simulation Model (PRISM) is an SD model developed by Jack Homer and collaborators to assist in planning and evaluating health care intervention strategies [38]. It represents major health conditions, including smoking, that contribute to cardiovascular disease. The population is divided according to sex and three age groups. Risk factors for the population are based on the Framingham Heart Study [39].

Advantages of our Approach and its Application

The models described above have aided our understanding of the population health effects of tobacco use and tobacco control policies. We have incorporated useful features from many of them in our modeling approach. Specifically, our model is an agent-based discrete-event model of population structure that uses a Markov chain

approach to describe state transitions due to tobacco use, while incorporating elements of dynamical systems and microsimulation modeling. Our approach can incorporate expected changes in demographics, behaviors, disease risks and the availability of different tobacco products; it allows flexibility to aggregate and disaggregate results for the national population and sub-populations; and it can represent expected stochastic variability due to uncertainties in demographic inputs and mortality rates. Our model also allows for the possibility of adding inter-agent communications and dynamics in the future. In this paper we demonstrate the efficacy of using hybridized modeling and an underlying behavioral approach for projecting future health states.

Methods

Population Structure Model (PSM) Conceptual Approach

The PSM represents a discrete-event, agent-based modeling approach offering support for microsimulation and cohort analysis of populations on the scale of specific small communities up to whole countries. Although it can be applied to a range of public health issues, in this paper, we model the population health effects of changing patterns of cigarette smoking. In the PSM, each individual is represented as a separate agent. Agents are characterized by intrinsic values, behavioral states and health states. Intrinsic values such as sex are permanent over the length of the simulation while other intrinsic values such as age change at each timestep. Behavioral states refer to the agent's available options at each timestep, such as initiating or quitting cigarette smoking, and health states characterize the possible health status of the individual at each timestep in the simulation.

An initial agent population is established based on empirical data and new agents are generated through births and immigration. Markov chains that capture the probability of state changes are used to represent agent behavior and consequential outcomes including deaths attributable to the behavior. The output results of the PSM allow us to project population health effects.

Figure 1 shows a conceptual diagram of the population flux (beginning in upper left) and the progress of a single sample agent through the course of a hypothetical lifetime using a Markov chain.

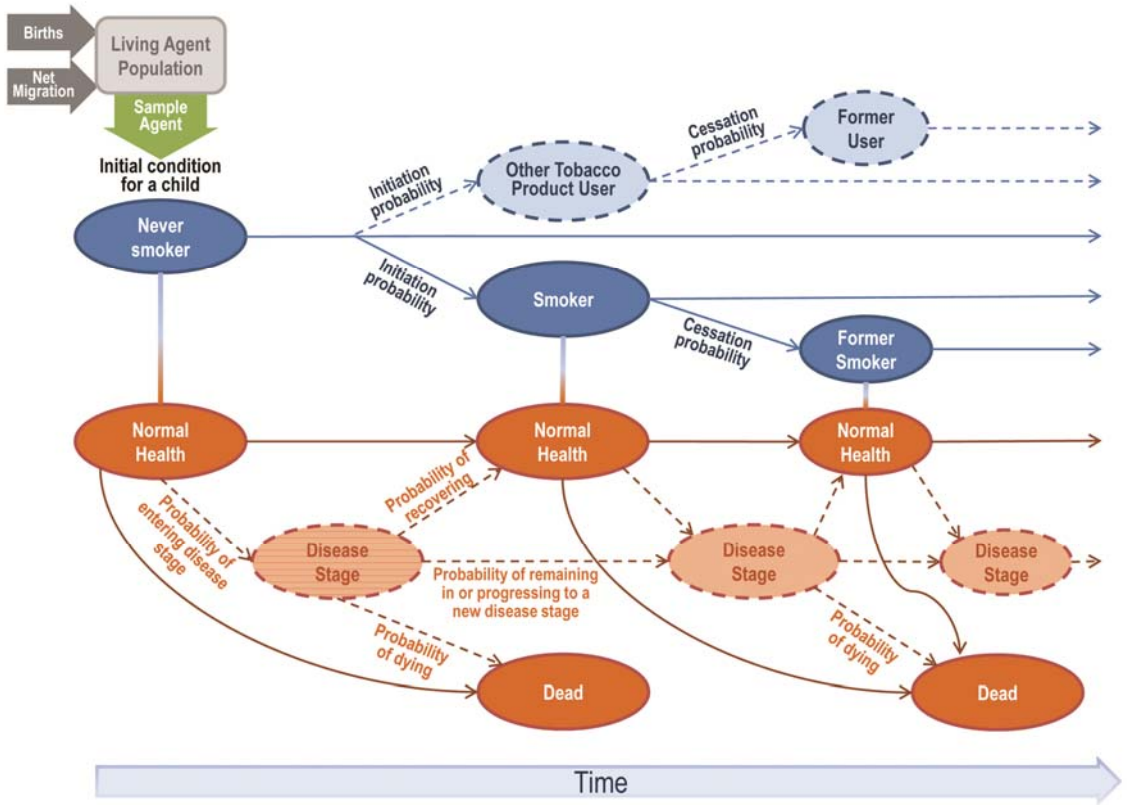


Figure 1. Conceptual Diagram for Population Structure Model (PSM) with Health and Smoking Transitions. Behavioral (blue) and health (orange) state transition probabilities depend on age, sex and smoking status Faded colors and dashed lines indicate potential elements that may be included in future versions of the PSM The agent population is modified through births, net migration, and deaths (upper left). In the transition probability example shown, a single agent's initial condition is that of a never-smoking child in normal health. The possible paths this agent may take as he or she ages are represented as arrows within a Markov chain, in which movement from one state to another is described by a transition probability. The probabilities associated with changes in smoking status (initiation and cessation rates) and health state vary by age, sex, and smoking status. At each timestep, the smoking status and health state of the agent are updated, while the model tracks each agent's age, smoking status (including time since cessation) and mortality. The figure does not illustrate all possible agent states.

The PSM uses agent-based modeling to represent individual behaviors, which are then characterized using Markov chains in a similar manner to work done by Sonnenberg and Beck [40] and Killeen [41]. As in microsimulation, PSM parameters are determined using empirical data available through sources such as the US Census and the Centers for Disease Control and Prevention (CDC). Although this version of the PSM does not model inter-agent interactions or small agent groupings (as is done in many microsimulation models), these features could be implemented in future versions. Finally, the PSM uses discrete-event modeling to manage updates of both dynamic system equations (such as birth and immigration) and agent-based state updates. Advantages of discrete-event modeling include simple model representation and operation as well as direct comparisons of model output to published data and estimates (using aligned population groups and timestep intervals that match published data).

Model Behavior Representation using Markov Chains

Changes in agent state are represented in our model using Markov chain transition probabilities. Markov chains are useful for representing states and probabilistic changes between states as shown in Figure 2.

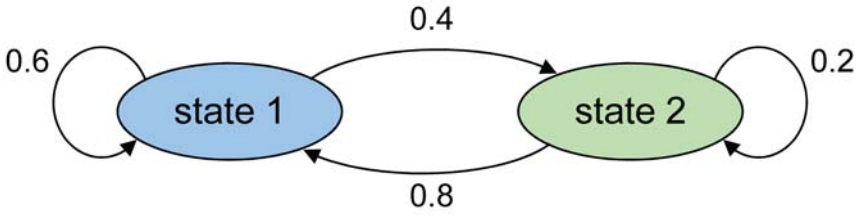


Figure 2. An Example Markov Chain. The diagram shows an example of a 2-state Markov chain where the probability of transitioning from state 1 to state 2 is 0.4, and the probability of transitioning from state 2 to state 1 is 0.8.

For this analysis, the states and transitions of interest include mortality status (i.e., "alive" or "dead") and smoking status (i.e., "never smoker," "current smoker" or "former smoker") as illustrated by the Markov chain in Figure 3.

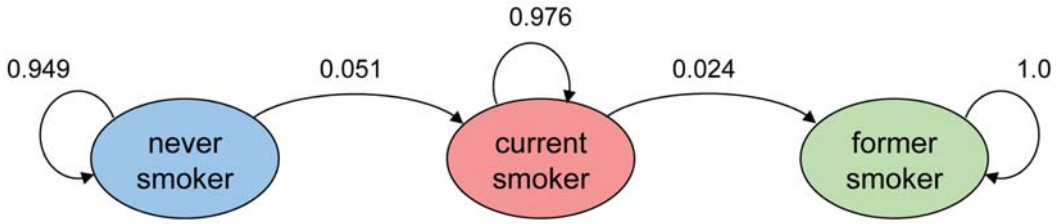


Figure 3. Example Markov Chain for an 18 Year-old Male's Cigarette Smoking Behavior. This Markov chain illustrates a specific example using transition probabilities directly from the PSM, developed from national health survey data.

In our model, long-term abstinence rates are used to model cessation, which incorporate quit attempts and relapse back to smoking. As described below, cessation rates are derived from cohort analyses of serial cross-sectional surveys [42]. Although the process of smoking cessation is often characterized by multiple quit attempts and failures, we incorporate the net impact of this cycling on long-term, successful cessation. This is demonstrated by the relatively low cessation rates used in the model (see Figure 5).

Implementation and Parameterization

Parameter Types and Data Sources

The bulleted listing below presents the data sources used to characterize the initial agent population and their transition probabilities. Population and health

characteristics from the year 2000, or as close as possible, were used to facilitate comparison of our projections with observed data from the last 10+ years.

The model is typically run using populations of 28,142 agents, 1/10,000th of the US population in the year 2000. It has also been run using a population of 2.8 million agents, a 1/100 scale. For output metrics involving large proportions of the population, such as future adult smoking prevalence, 28,142 is an adequate agent population; larger numbers of agents are required to obtain reasonable resolution when examining impacts to smaller groups, such as when calculating smoking-attributable deaths within a limited age group.

Data sources and usage for parameterization with application to cigarette smoking are given below.

- Initial Population
    - Population counts by sex and age group come from the 2000 US Census [43].
- Changes to Population: Future births
    - Annual births are proportional to the US female population of reproductive ages. The proportionality constant was determined from US Census birth projections released in 2008 [43], [44].
- Changes to Population: Future deaths
    - Underlying age- and sex-specific never-smoker death rates for ages 35 and over for year 2000 are obtained from National Health Interview Survey (NHIS) data linked for mortality follow-up [45].
    - Age- and sex-specific deaths rates for all persons under 35 years of age are obtained from U.S. vital statistics data [46]. Given that no excess mortality risk from smoking is applied prior to age 35, these rates are the same for never, former, and current smokers for these ages (35 and under).
    - Death rates for current and former smokers are calculated by multiplying these underlying never-smoker death rates by relative risks according to sex, age, smoking status, and, for former smokers, years since cessation. Relative risk data come from the American Cancer Society Cancer Prevention Study II (CPS-II) [47]-[49] and a smoothing function was applied.
    - For former smokers, a lag period of two years is incorporated after smoking cessation before a reduction in mortality risk is applied. This corresponds to the lag period between assessment of smoking status and mortality follow-up in the CPS-II relative risks.
    - Scale factors obtained using the Lee-Carter mortality projection method [50] are used to scale the underlying never-smoker death rates over time in order to incorporate future expected decreases in mortality.

- Changes to Population: Future immigration
    - Annual numbers of immigrants are proportional to the US population. The proportionality constant was determined from US Census net international migration projections released in 2008 [51].
- Initial Smoking Status
    - For adults ages 18 and over: Current and former smoker prevalence by age, sex, and time since quitting for the US in 2000 are obtained from NHIS data [52].
    - For youth less than age 18: Current and former smoker prevalence by age and sex are obtained from reconstructions of smoking prevalence for historical birth cohorts from national health survey (NHIS) data [53].
- Changes to Smoking Status: Each year, agents have age-specific probabilities of transitioning from never smokers to current smokers (for ages 8 to 30) and from current smokers to former smokers. Former smokers are not allowed to become current smokers again.
    - US Population
        - Age- and sex-specific initiation probabilities, for ages 30 and under, and cessation probabilities, for all ages, are obtained from reconstructions of smoking prevalence for historical birth cohorts from national health survey (NHIS) data [53]. After age 30 initiation probabilities are negligible and are therefore assumed to be 0.
        - Probabilities remain constant into the future for the model scenarios described here.
    - New Immigrants:
        - Smoking prevalence among new immigrants comes from 2007-11 NHIS estimates [52]. Values used are 19.1% smoking prevalence for men and 4.4% for women. Former smoking prevalences are 13.1% for men and 5.6% for women. Values for smoking prevalence for new immigrants remain constant over time.
        - After immigration, immigrants have the same probabilities for initiation and cessation as the general population.

Generating an Initial Population

US Census and National Health Interview Survey data are used to estimate US population proportions in 2000 by sex, age group, and smoking status, shown in Figure 4. These data are shown as proportions of the total population but are used as a probability distribution to stochastically generate an initial population that is representative of the US in terms of sex, age, and smoking status. This is useful in modeling the US population because data inputs from nationally representative

samples, such as the National Health Interview Survey, have some inherent uncertainty.

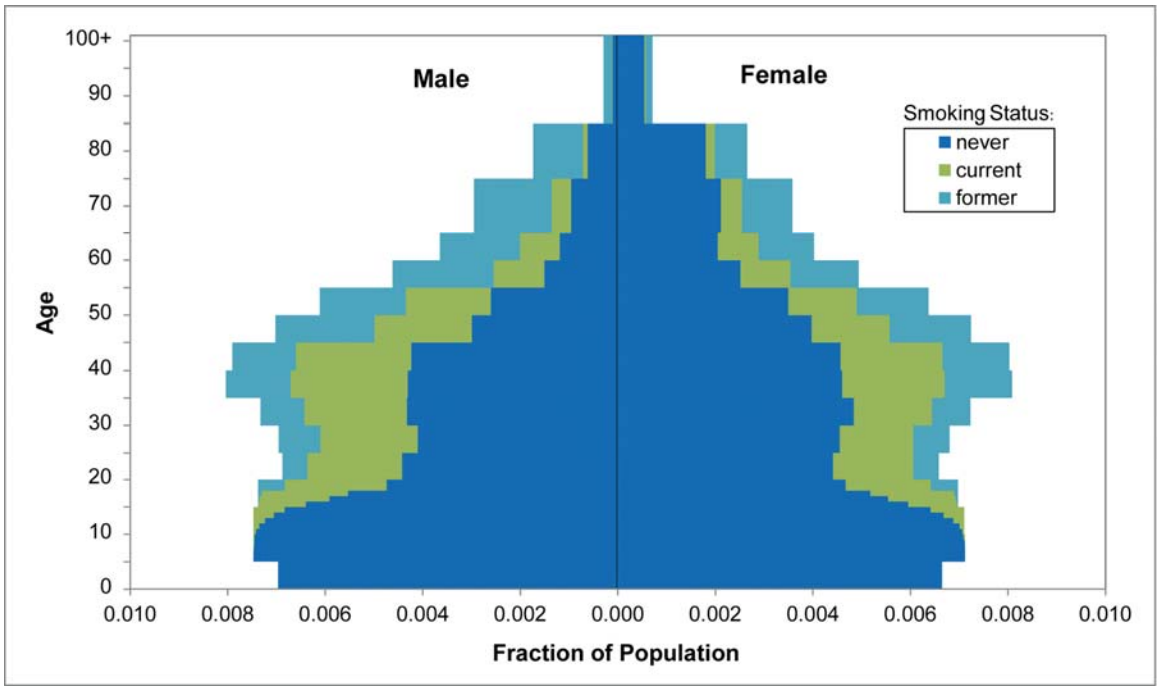


Figure 2. Initial Model Population Distribution based on the 2000 Census by Age, Sex, and Smoking Status.

Birth and Immigration

Annual births are proportional to the US female population of reproductive ages. Because smoking-related mortality does not occur in the model until age 35, reproductive ages in the model are restricted to 15-34. This allows us to keep calculations of smoking-related deaths and birth rate decoupled, so that an equivalent number of new births are modeled when comparing different scenarios of smoking initiation and cessation.

Consistent with the US Census, annual numbers of immigrants are proportional to the entire US population.

Computing the Effects of Cigarette Smoking

Cigarette smoking initiation and cessation rates are obtained from reconstructions of historical birth cohorts derived from multiple years of the NHIS (see Anderson et al., 2012 for detailed methodology). The age- and sex-specific initiation and cessation rates used in the model are shown in Figures 5a and 5b. The jagged nature of the curves reflects the combination of multiple cohorts. In the smoking prevalence projections modeled in this paper, initiation and cessation rates are assumed to remain constant in the future.

Age- and sex-specific death rates for never smokers ages 35 and over for year 2000 are obtained from NHIS data linked for mortality follow-up [45]. Age- and sex-specific deaths rates for all persons under 35 years of age are obtained from U.S. vital statistics data [46]. Given that no excess mortality risk from smoking is applied prior to age 35, these rates are the same for never, former, and current smokers.

Mortality rates for current and former smokers are obtained by multiplying smoking relative risks by never smoker mortality rates. An individual's probability of dying in a given year is given in Equation (2).

$$\Pr\{\text{death}\}_i = RR_i * \Pr\{\text{death} \mid \text{never-smoker}\}_i \quad (2)$$

In Equation (2), the probability of death for a never smoker is a function of the individual's sex and age. The relative risk (*RR*) associated with current or former cigarette smoking is a function of the individual's age, sex, smoking status, and, for former smokers, years since cessation.

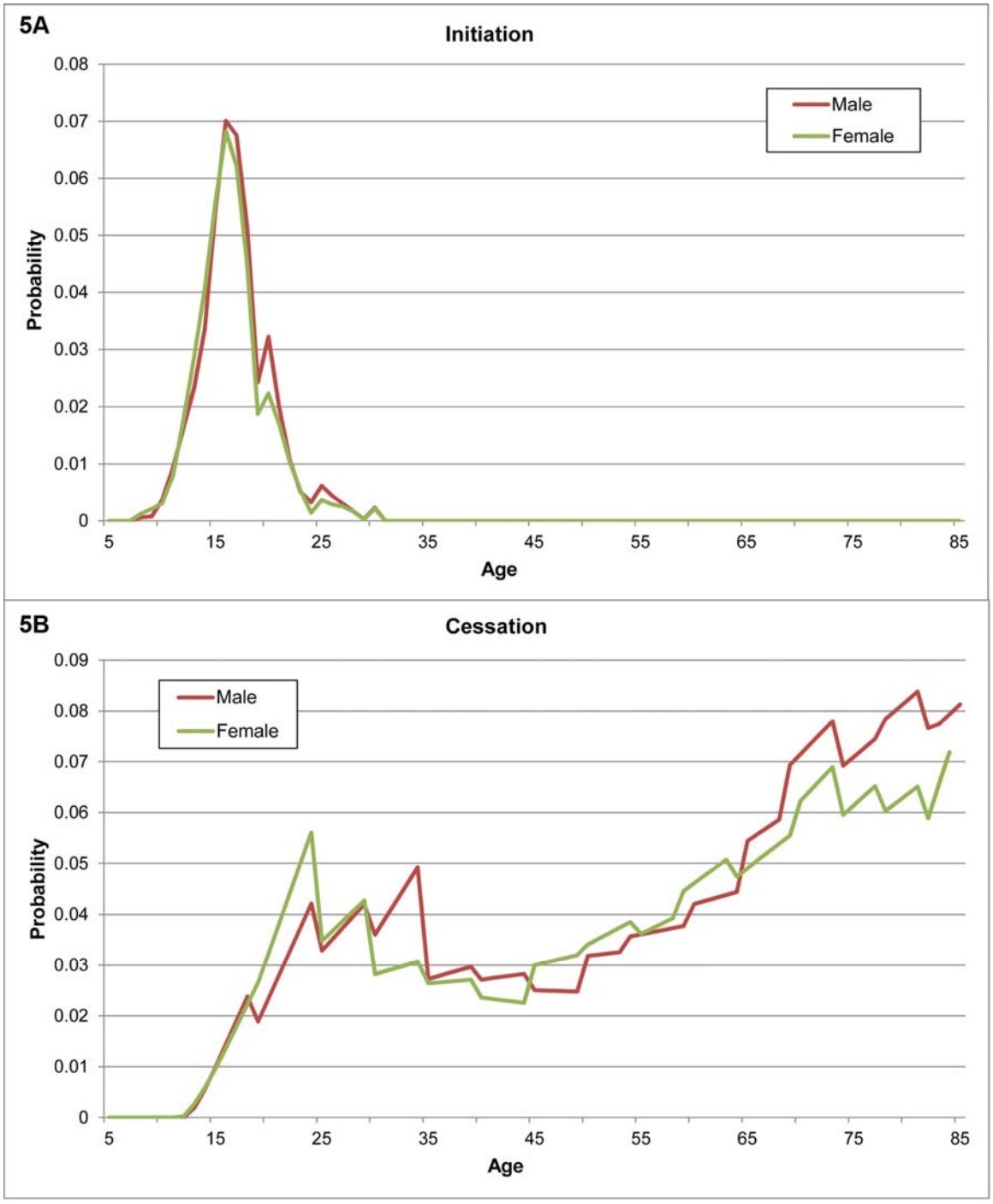


Figure 5. Annual Smoking Initiation (5a) and Cessation (5b) Rates by Age and Sex. Input data used for both initiation and cessation probabilities come from analysis of multiple cohorts from NHIS data, which accounts for the jagged nature of these input data.

Simulations

To examine future cigarette smoking prevalence for adults ages 18 and over, we simulated a scenario in which smoking initiation and cessation rates are kept constant throughout the simulation from 2000 to 2050. Our initial population characteristics (age, sex, and smoking status) are generated stochastically: 100 simulations were conducted, each utilizing a different random seed, resulting in a slightly different initial population and behavior throughout the simulation. Each initial population is comprised of 28,142 agents ($1/10{,}000^{th}$ of the year 2000 US population). The stochastic variability introduced by the model's initial population and health and behavioral transition probabilities represents the expected variability of the projections, given the estimated input values. We report the average smoking prevalence across the 100 runs, along with 95% confidence intervals.

The PSM also projects smoking-attributable mortality for the US over time. Smoking-attributable deaths for a particular year are estimated by comparing a scenario in which current and former smokers' relative risks (RRs) are applied normally, with one in which all current and former smoker relative risks are set to 1.0. The difference between the numbers of deaths in these two scenarios is the number of deaths attributable to smoking in the US in that year. We present the average of 100 simulation runs, along with the 95% confidence interval.

Results

Model Validation

Population Projection

To validate the model, we first compared PSM-generated US population projections to those of the US Census Bureau [54]. Our initial population projections, shown in red in Figure 6, were lower than those of the Census Bureau. This discrepancy was addressed by including time-varying mortality scale factors derived from the Lee-Carter mortality forecasting method [50] that account for projected future decreases in mortality rates, as shown in Figure 7. The resulting population projections are highly consistent with Census projections.

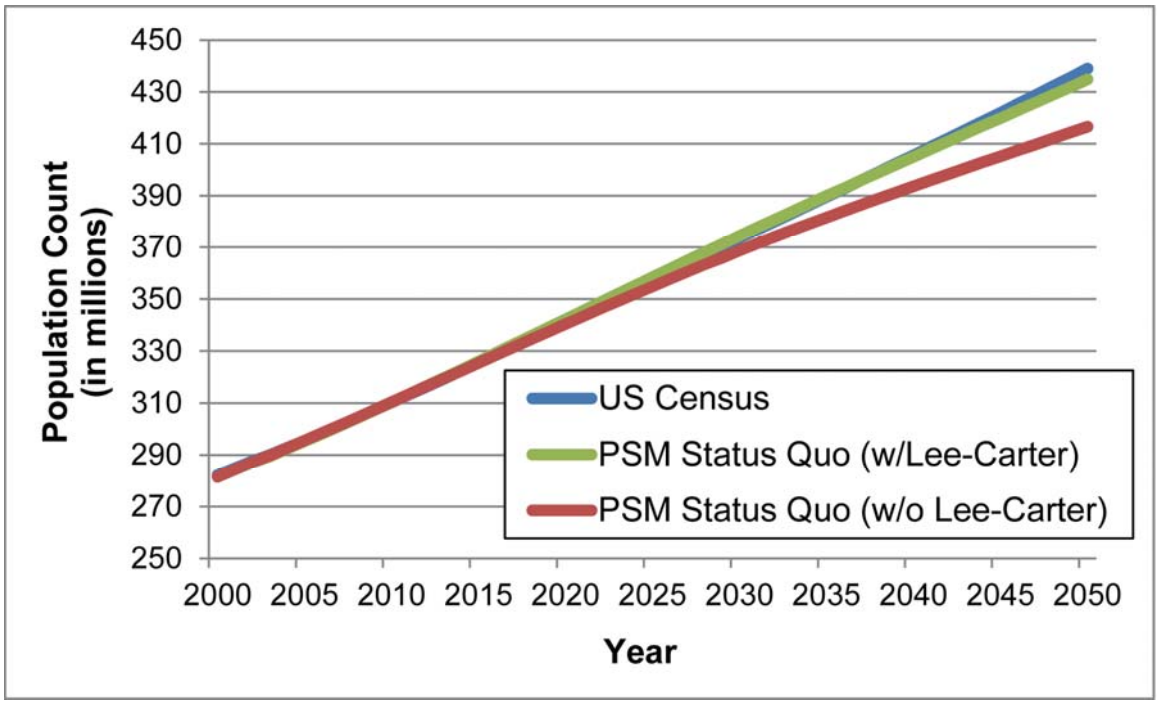


Figure 6. PSM Population Projection with and without Lee-Carter Mortality Forecasting Method Compared with US Census Projection.

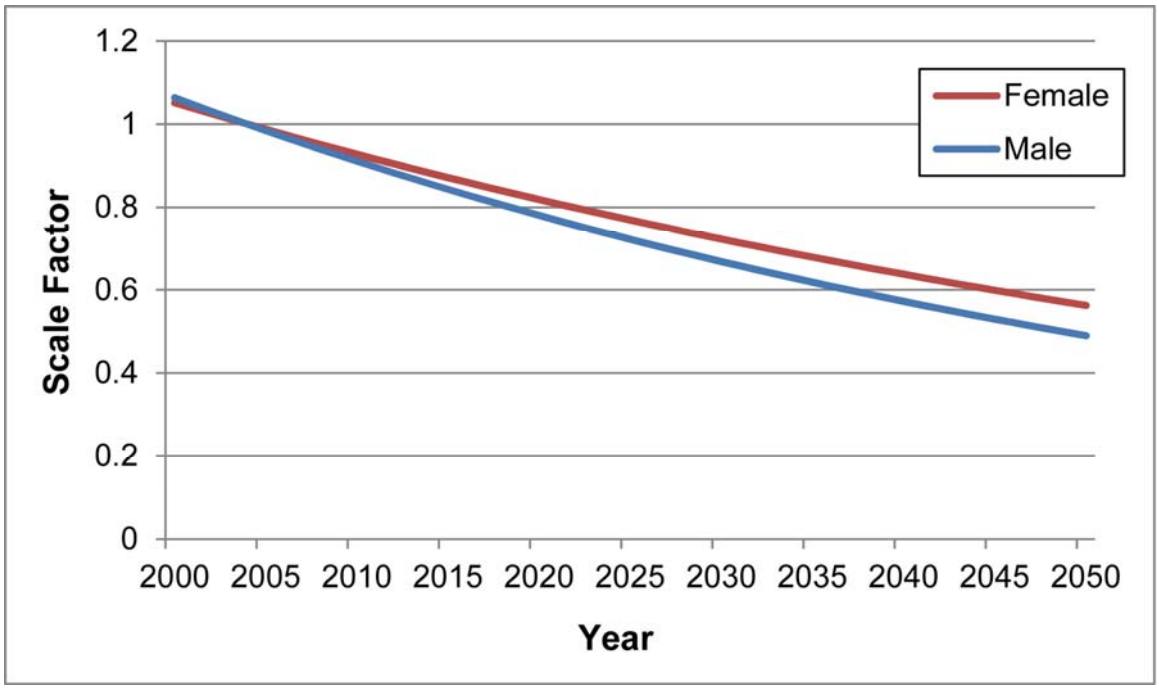


Figure 7. Lee-Carter Mortality Scale Factors for Age 50 by Sex.

Population projections from our model can be also disaggregated into subgroups using characteristics such as age group and sex. For example, we project a US

population count of 215 million males and 220 million females in 2050, consistent with US Census projections of 216.5 million and 223.9 million males and females, respectively.

Adult Smoking Prevalence

Figure 8 presents the projected cigarette smoking prevalence for adults ages 18 and over, under a status quo scenario in which smoking initiation and cessation rates are kept constant throughout the simulation from year 2000 to 2050. Figure 8 also compares projections from our model, labeled "PSM Status Quo," with CDC estimates [55] and the results of other published modeling analyses [35], [36], [56]. The results show consistent agreement.

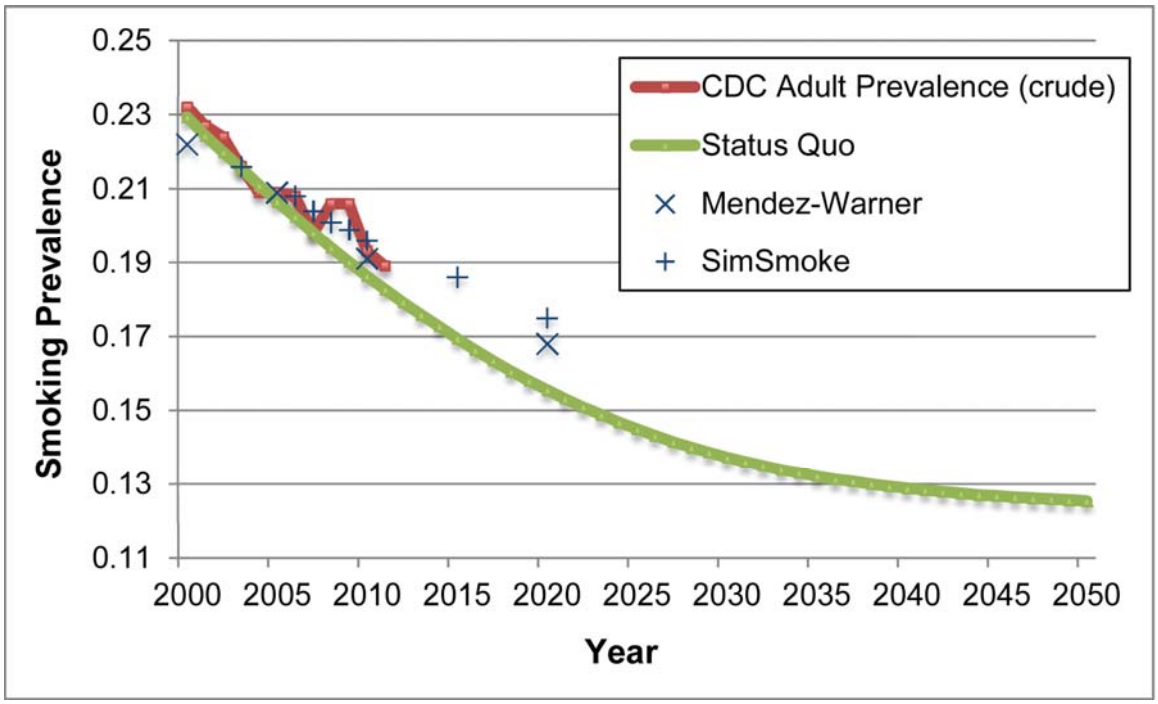


Figure 8. Comparison of Adult Smoking Prevalence in Population Models and CDC Data. PSM Status Quo results are the average of 100 simulation runs. PSM adult prevalence projections compare favorably with results from the SimSmoke and Mendez-Warner models described above. Results from Mendez-Warner for 2000 and 2005 are from their 2000 paper [35], and 2010 and 2020 results are from their 2008 paper [36]. Results from SimSmoke are taken from a recent paper in which the SimSmoke model is initialized using 2003 CDC prevalence values [56]. The PSM average prevalence values reported for the 100 simulations are quite smooth; maximum values of the 95% confidence interval are 0.1% of the calculated mean.

All of the models show a similar downward trend in smoking prevalence. Our status quo PSM scenario, which uses constant sex- and age-specific initiation and cessation rates, projects declines in smoking prevalence over time due to smoking initiation rates having fallen in decades prior to 2000. Smoking prevalence approaches a steady state and reaches a value of about 12.5% by 2050. Figure 9 shows PSM projections of adult current and former smoking prevalence by sex, along with comparisons to CDC estimates.

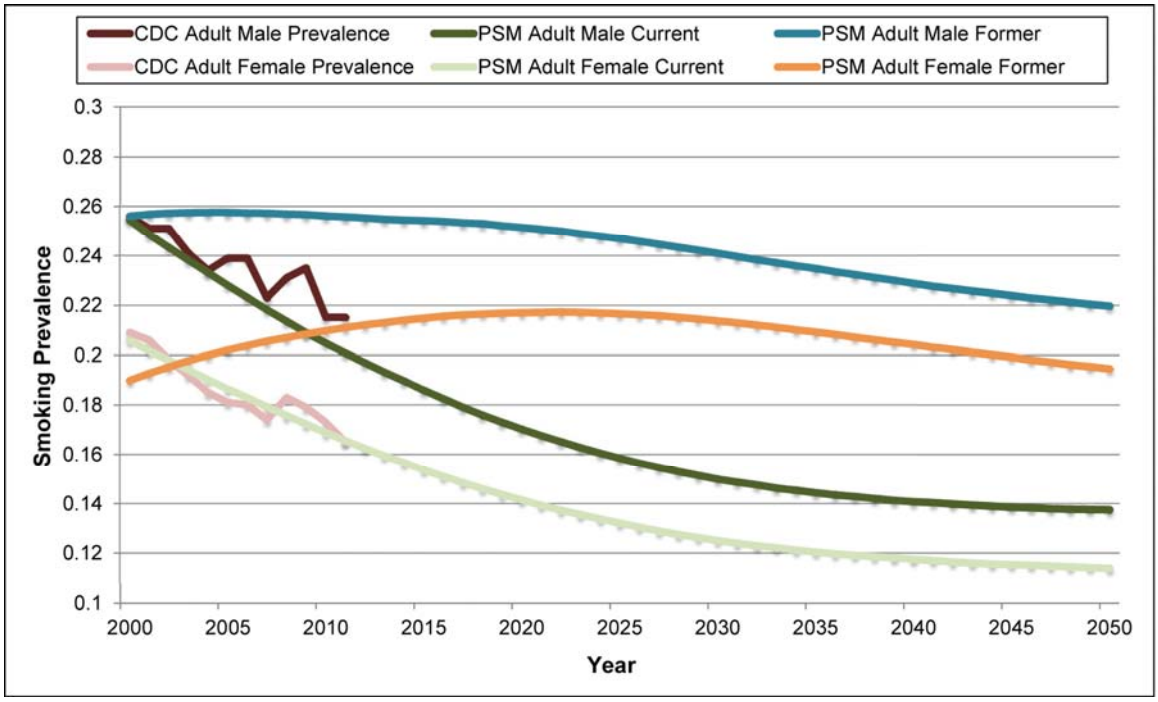
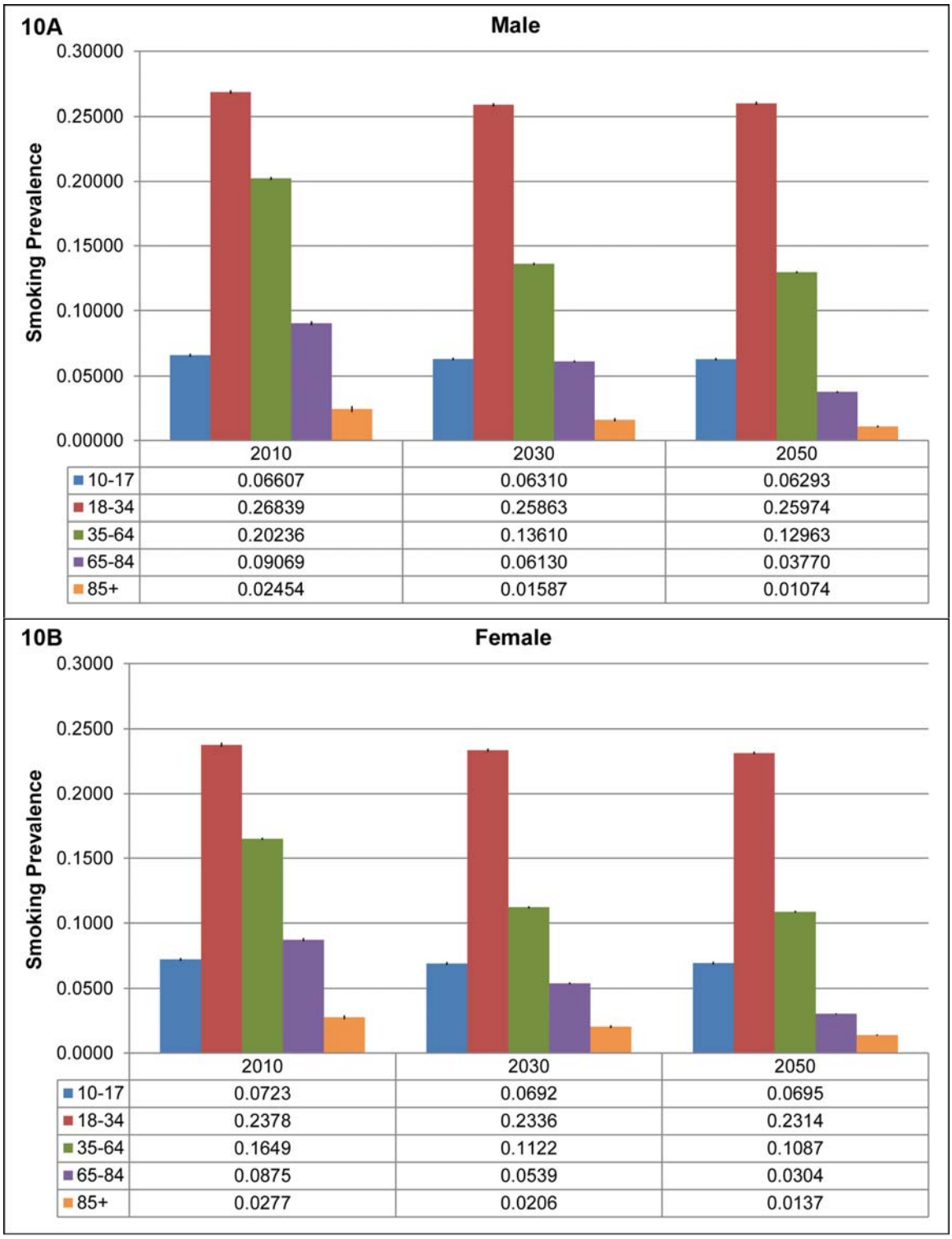


Figure 9. PSM Adult Smoking Prevalence Projections and CDC Estimates by Sex.

Figures 10a and 10b present PSM prevalence projections by age group and sex. Although youth smoking prevalence is typically reported as smoking in the past 30 days, our estimates among youth reflect more frequent, sustained use. The 10-17 and 18-34 year old age groups are very close to their equilibrium prevalence by 2010 as a result of initiation and cessation rates being held constant in the simulation beginning in 2000. Older age groups show decreases in prevalence over time as these groups become populated by individuals who have experienced lower initiation and higher cessation rates.



	2010	2030	2050
10-17	0.06607	0.06310	0.06293
18-34	0.26839	0.25863	0.25974
35-64	0.20236	0.13610	0.12963
65-84	0.09069	0.06130	0.03770
85+	0.02454	0.01587	0.01074

	2010	2030	2050
10-17	0.0723	0.0692	0.0695
18-34	0.2378	0.2336	0.2314
35-64	0.1649	0.1122	0.1087
65-84	0.0875	0.0539	0.0304
85+	0.0277	0.0206	0.0137

Figure 10a and 10b. Projected male (10a) and female (10b) smoking prevalence by age group and time. Error bars represent 95% confidence intervals across 100 runs.

Smoking-Attributable Deaths

For the year 2000, using all-cause mortality relative risks, the PSM computes the 401,000 smoking-attributable deaths for ages 35 and over with a 95% confidence interval (CI) of 389,000-413,000. This estimate is consistent with the annual CDC estimate for 2000-2004 of 393,000 (not including second-hand smoke and in-home fire deaths) and the SimSmoke reported value of 418,317 for 2000 [30].

Conclusions/Future Work

We have presented a population structure model capable of forecasting the effects of changes in individual behaviors on overall population health. We have used cigarette smoking as an example to demonstrate the value of our approach in providing projections that can be used for policy formulation and evaluation. This approach can be easily adapted to other risk behaviors and conditions that affect health at the population level, such as alcohol consumption and obesity.

In this example, our model uses a 1:10,000 scaled population that is nationally representative of the US in terms of demographic characteristics and smoking status. We established a baseline status quo projection using constant sex- and age-specific smoking initiation and cessation rates that allows for comparison with scenarios in which tobacco control measures affect these rates. Our model projects future smoking prevalence and smoking-attributable mortality. With appropriate inputs, the model can be easily extended to compute estimates of cause-specific mortality, morbidity, quality-adjusted life expectancy, and health care costs. The model can be applied to other types of tobacco products, and additional demographic characteristics such as race/ethnicity can be easily incorporated because of the agent-based approach. Model parameters can be represented by probability distributions, thus allowing for the estimation of variability for results and projections.

We have shown that our model produces population and health projections that are consistent with published data and estimates. Our use of time-varying demographic inputs, such as properly scaled underlying never smoker mortality rates, produces population projections that are consistent with US Census projections. These techniques will allow us to project future health outcomes such as smoking-attributable mortality with greater accuracy than has been possible previously.

Finally, in designing and implementing this model, we have shown that a hybrid discrete-event agent-based model has many advantages in projecting population and health outcomes compared with any other single type of modeling approach. This modeling approach generates agents with specific demographic and smoking characteristics, allowing us to characterize the behavior of each agent at the individual level. We can aggregate results for the entire population or for demographic subpopulations of particular interest. Expansion of this model, as applied to tobacco, will include multiple products, cause-specific morbidity and mortality, and more specificity in demographic characteristics. These extensions

will allow us to model even more fully the possible effects of patterns in tobacco use on future morbidity and mortality in the US.

Acknowledgments:

The authors would like to thank Corinne Husten and Robert J. Glass for providing overall guidance and direction. Contributions were made to the technical direction of the work by Bridget Ambrose and Shawn Fultz. Aldo Zagonel initially explored a system dynamics approach for this problem and provided valuable assistance in framing the problem and identifying modeling and data needs. Thomas Corbet, Jr. contributed to the original discrete-event study design. Louise Maffitt provided technical writing assistance and constructive review. The authors would like to thank David Burns and Michael Thun for providing estimates for initiation and cessation rates and mortality relative risk, respectively.

References

1. White GC (2000) Modeling Population Dynamics. In: Demarais S, Krausman PR, editors. Ecology and Management of Large Mammals in North America. Upper Saddle River, New Jersey, USA: Prentice-Hall. pp. 84-107.
2. Turchin P (2003) Complex Population Dynamics: a Theoretical/Empirical Synthesis. Princeton, NJ: Princeton University Press.
3. de Boer R (2011) Modeling Population Dynamics: a Graphical Approach. Utrecht University. http://theory.bio.uu.nl/rdb/books/mpd.pdf.
4. McIntosh R (1985) The Background of Ecology: Cambridge University Press.
5. Renshaw E (1991) Modeling Biological Populations in Space and Time: Cambridge University Press.
6. Gilbert N, Troitzsch KG (2005) Simulation for the Social Scientist. Berkshire: Open University Press.
7. Borshchev A, Filippov A (2004) From System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Techniques, Tools. In: Kennedy M, Winch GW, Langer RS, Rowe JI, Yanni JM, editors; Oxford.
8. Allen T (2011) Introduction to Discrete Event Simulation and Agent-based Modeling. London: Springer.
9. Berryman AA (1992) The Origins and Evolution of Predator-Prey Theory. Ecology 73: 1530-1535.
10. Forrester JW (1980) Principles of Systems. Cambridge, MA: MIT Press.
11. Sterman JD (2001) System dynamics modeling: Tools for learning in a complex world. California Management Review 43: 8-25.

12. Horner J, Hirsch G (2006) System Dynamics Modeling for Public Health: Background and Opportunities. American Journal of Public health 96: 452-458.

13. Tobias MI, Cavana RY, Bloomfield A (2010) Application of a System Dynamics Model to Inform Investment in Smoking Cessation Services in New Zealand. American Journal of Public Health 100: 1274-1281.

14. Billari F, Prskawetz A (2005) Studying population dynamics from the bottom-up: the crucial role of agent-based computational demography. XXV IUSSP International Population Conference. http://demoscope.ru/weekly/knigi/tours_2005/papers/iussp2005s50673.pdf.

15. Bonabeau E (2002) Agent-based modeling: Methods and techniques for simulating human systems. Proceedings of the National Academy of Sciences 99: 7280-7287.

16. Zhang Q, Jager W (2011) Agent-based modeling of population dynamics in municipalities: Migration in the Derbyshire & Nottinghamshire cases in the UK. Report for the EU Prima project, WP3. University of Groningen, Center for Social Complexity Studies.

17. (1986) Microanalytic simulation models to support social and financial policy; Orcutt G, Merz J, Quincke H, editors. Amsterdam: Elsevier (North Holland Publishing Co.).

18. (1996) Microsimulation and Public Policy; Harding A, editor. Amsterdam: Elsevier.

19. Fone D, Hollinghurst S, Temple M, Round A, Lester N, et al. (2003) Systematic review of the use and value of computer simulation modeling in population health and health care delivery. Journal of Public Health Medicine 25: 325-335.

20. Warner KE, Smith RJ, Smith DG, Fries BE (1996) Health and economic implications of a work-site smoking-cessation program: a simulation analysis. Journal of Occupational and Environmental Medicine 38: 981-992.

21. Richardson GP (2007) How to Anticipate Change in Tobacco Control Systems. In: Best A, Clark PI, Leischow SJ, Trochim WMK, editors. Greater Than the Sum: Systems Thinking in Tobacco Control: U.S. Department of Health and Human Services. pp. 109-146.

22. Roberts EB, Homer J, Kasabian A, Varrel M (1982) A systems view of the smoking problem. International Journal of Biomedical Computing 13: 69-86.

23. Cavana RY, Tobias M (2008) Integrative system dynamics: analysis of policy options for tobacco control in New Zealand. Systems Research & Behavioral Science 25: 675-694.

24. Tengs TO, Osgood ND, Lin TH (2001) Public health impact of changes in smoking behavior: results from the Tobacco Policy Model. Medical Care 39: 1131-1141.

25. Tengs TO, Osgood ND, Chen LL (2001) The cost-effectiveness of intensive national school-based anti-tobacco education: results from the tobacco policy model. Preventive Medicine 33: 558-570.

26. Apelberg BJ, Onicescu G, Avila-Tang E, Samet JM (2010) Estimating the Risks and Benefits of Nicotine Replacement Therapy for Smoking Cessation in the United States. American Journal of Public Health 100: 341-348.

27. Levy DT, Cummings KM, Hyland A (2000) A simulation of the effects of youth initiation policies on overall cigarette usage. American Journal of Public Health 90: 1311-1314.

28. Levy DT, Cummings KM, Hyland A (2000) Increasing taxes as a strategy to reduce cigarette use and deaths: results of a simulation model. Preventive Medicine 31: 279-286.

29. Levy DT, Friend K (2002) A simulation model of policies directed at treating tobacco use and dependence. Medical Decision Making 22: 6-17.

30. Levy DT, Friend K (2002) Examining the effects of tobacco treatment policies on smoking rates and smoking related deaths using the SimSmoke computer simulation model. Tobacco Control 11: 47-54.

31. Levy DT, Nikolayev L, Mumford E (2005) Recent trends in smoking and the role of public policies: results from the SimSmoke tobacco control policy simulation model. Addiction 100: 1526-1536.

32. Levy DT, Bales S, Nguyen TL, Nikolayev L (2006) The role of public policies in reducing smoking and deaths caused by smoking in Vietnam: Results from the Vietnam tobacco policy simulation model. Social Science & Medicine 62: 1819-1830.

33. Nagelhout GE, Levy DT, Blackman K, Currie L, Clancy L, et al. (2012) The effect of tobacco control policies on smoking prevalence and smoking-attributable deaths. Findings from the Netherlands SimSmoke Tobacco Control Policy Simulation Model. Addiction 107: 407-416.

34. Mendez D, Warner KE, Courant PN (1998) Has smoking cessation ceased? Expected trends in the prevalence of smoking in the United States. American Journal of Epidemiology 148: 249-258.

35. Mendez D, Warner KE (2000) Smoking prevalence in 2010: Why the Healthy People goal is unattainable. American Journal of Public Health 90: 401-403.

36. Mendez D, Warner KE (2008) Setting a challenging, yet realistic smoking prevalence target for Healthy People 2020: Learning from the California experience. American Journal of Public Health 98: 556-559.

37. Mendez D (2011) Results from a Population Dynamics Model of the Consequences of Menthol Cigarettes for Smoking Prevalence and Disease Risks. Appendix A of the Tobacco Products Science Advisory Committee Menthol Report.

38. Homer J, Milstein B, Wile K, Trogdon J, Huang P, et al. (2010) Simulating and evaluating local interventions to improve cardiovascular health. Preventing Chronic Disease 7. http://www.cdc.gov/pcd/issues/2010/jan/08_0231.htm.

39. Anderson KM, Odell PM, Wilson PW, Kannel WB (1991) Cardiovascular disease risk profiles. American Heart Journal 121: 293-298.

40. Sonnenberg FA, Beck JR (1993) Markov models in decision making: A practical guide. Medical Decision Making 13: 322-338.

41. Killeen PR (2011) Markov model of smoking cessation. To appear in Proceedings of the National Academy of Sciences, available online (for early release). http://www.pnas.org/content/early/2011/04/19/1011277108.

42. US Census (2011) Table 7. Resident Population by Sex and Age: 1980 to 2009. http://www.census.gov/compendia/statab/2011/tables/11s0007.pdf.

43. US Census (2008a). Projected Births by Sex, Race, and Hispanic Origin for the United States: July 1, 2000 to June 30, 2050. http://www.census.gov/population/www/projections/downloadablefiles.html.

44. Martin JA, Hamilton BE, Sutton PD, Ventura SJ, Matthew TJ, et al. (2010) Births: Final Data for 2007. National Vital Statistics Reports 58. http://www.cdc.gov/nchs/data/nvsr/nvsr58/nvsr58_24.pdf.

45. (2010) National Center for Health Statistics. National Health Interview Survey Linked Mortality Files (NHIS-LMF). http://www.cdc.gov/nchs/data_access/data_linkage/mortality/nhis_linkage.htm.

46. Miniño AM, Arias E, Kochanek KD, Murphy SL, Smith BL (2002) Deaths: Final Data for 2000. National Vital Statistics Reports 50. http://www.cdc.gov/nchs/data/nvsr/nvsr50/nvsr50_15.pdf.

47. Thun MJ, Day-Lally C, Myers DG, Calle EE, Flanders WD, et al. (1997) Trends in Tobacco Smoking and Mortality From Cigarette Use in Cancer Prevention Studies 1 (1959 Through 1965) and II (1982 Through 1988). Smoking and Tobacco Control Monograph, 8. pp. 305-382. http://cancercontrol.cancer.gov/tcrb/monographs/8/index.html.

48. Thun MJ, Myers DG, Day-Lally C, Namboordiri MM, Calle EE, et al. (1997) Age and the Exposure-Response Relationships Between Cigarette Smoking and Premature Death in Cancer Prevention Study II. Smoking and Tobacco Control Monograph, 8. pp. 383-475. http://cancercontrol.cancer.gov/tcrb/monographs/8/index.html.

49. Thun MJ (2010) Personal communication.

50. Lee RD, Carter LR (1992) Modeling and Forecasting U.S. Mortality. Journal of the American Statistical Association 87: 659-671. http://www.jstor.org/stable/10.2307/2290201.

51. US Census (2008b). Projected Net International Migration by Sex, Race, and Hispanic Origin for the United States: July 1, 2000 to June 30, 2050. http://www.census.gov/population/www/projections/downloadablefiles.html.

52. (2012) National Center for Health Statistics. National Health Interview Survey (NHIS). http://www.cdc.gov/nchs/nhis.htm.

53. Anderson CM, Burns DM, Dodd KW, Feuer EJ (2012) Birth cohort specific estimates of smoking behaviors for the U.S. population. Risk Analysis 32: S14-S24. http://cisnet.cancer.gov/resources/impact_reviewID_A3Z.html.

54. US Census (2011) Table 3. Resident Population Projections: 2010 to 2050. http://www.census.gov/compendia/statab/2011/tables/11s0007.pdf.

55. Centers for Disease Control and Prevention, CDC (2011) Early Release of Selected Estimates Based on Data From the 2010 National Health Interview Survey. http://www.cdc.gov/nchs/data/nhis/earlyrelease/201106_08.pdf.

56. Levy DT, Mabry PL, Graham AL, Orleans CT, Abrams DB (2010) Reaching Healthy People 2010 by 2013: A SimSmoke Simulation. American Journal of Preventive Medicine 38: S373-S381.

FIGURE TITLES AND LEGENDS

Figure 3. Conceptual Diagram for Population Structure Model (PSM) with Health and Smoking Transitions

Behavioral (blue) and health (orange) state transition probabilities depend on age, sex and smoking status Faded colors and dashed lines indicate potential elements that may be included in future versions of the PSM The agent population is modified through births, net migration, and deaths (upper left). In the transition probability example shown, a single agent's initial condition is that of a never-smoking child in normal health. The possible paths this agent may take as he or she ages are represented as arrows within a Markov chain, in which movement from one state to another is described by a transition probability. The probabilities associated with changes in smoking status (initiation and cessation rates) and health state vary by age, sex, and smoking status. At each timestep, the smoking status and health state of the agent are updated, while the model tracks each agent's age, smoking status (including time since cessation) and mortality. The figure does not illustrate all possible agent states.

Figure 2. An Example Markov Chain

The diagram shows an example of a 2-state Markov chain where the probability of transitioning from state 1 to state 2 is 0.4, and the probability of transitioning from state 2 to state 1 is 0.8.

Figure 3. Example Markov Chain for an 18 Year-old Male's Cigarette Smoking Behavior

This Markov chain illustrates a specific example using transition probabilities directly from the PSM, developed from national health survey data.

Figure 4. Initial Model Population Distribution based on the 2000 Census by Age, Sex, and Smoking Status

Figure 5. Annual Smoking Initiation (5a) and Cessation (5b) Rates by Age and Sex

Input data used for both initiation and cessation probabilities come from analysis of multiple cohorts from NHIS data, which accounts for the jagged nature of these input data

Figure 6. PSM Population Projection with and without Lee-Carter Mortality Forecasting Method Compared with US Census Projection

Figure 7. Lee-Carter Mortality Scale Factors for Age 50 by Sex

Figure 8. Comparison of Adult Smoking Prevalence in Population Models and CDC Data

PSM Status Quo results are the average of 100 simulation runs. PSM adult prevalence projections compare favorably with results from the SimSmoke and Mendez-Warner models described above. Results from Mendez-Warner for 2000 and 2005 are from their 2000 paper [35], and 2010 and 2020 results are from their 2008 paper [36]. Results from SimSmoke are taken from a recent paper in which the SimSmoke model is initialized using 2003 CDC prevalence values [56]. The PSM average prevalence values reported for the 100 simulations are quite smooth; maximum values of the 95% confidence interval are 0.1% of the calculated mean.

Figure 9. PSM Adult Smoking Prevalence Projections and CDC Estimates by Sex

Figure 10a and 10b. Projected male (10a) and female (10b) smoking prevalence by age group and time

Error bars represent 95% confidence intervals across 100 runs.